(Article begins on next page)

Diffusion Estimation over Cooperative Networks with Missing Data

Mohammad Reza Gholami and Erik G. Ström
Department of Signals and Systems
Chalmers University of Technology, Gothenburg, SE-412 96, Sweden

Ali H. Sayed
Department of Electrical Engineering
University of California, Los Angeles, CA 90095

***Abstract*—In many fields, and especially in the medical and social sciences and in various recommender systems, data are often gathered through clinical studies or targeted surveys. Participants are generally reluctant to respond to all questions in a survey or they may lack information to respond adequately to the questions. The data collected from these studies tend to lead to linear regression models where the regression vectors are only known partially: some of their entries are either missing completely or replaced randomly by noisy values. There are also situations where it is not known beforehand which entries are missing or censored. There have been many useful studies in the literature on techniques to perform estimation and inference with missing data. In this work, we examine how a connected network of agents, with each one of them subjected to a stream of data with incomplete regression information, can cooperate with each other through local interactions to estimate the underlying model parameters in the presence of missing data. We explain how to modify traditional distributed strategies through regularization in order to eliminate the bias introduced by the incomplete model. We also examine the stability and performance of the resulting diffusion strategy and provide simulations in support of the findings. We consider two applications: one dealing with a mental health survey and the other dealing with a household consumption survey.**

I. INTRODUCTION

In data gathering procedures, it is common that some components of the data are missing or left unobserved. For example, in a clinical study, a participant may be reluctant to answer some questions. Likewise, in a recommender system using content based filtering [1], a participant may prefer to leave some questions unanswered. The phenomenon of missing data is ubiquitous in many fields including the social sciences, medical sciences, econometrics, and machine learning [2]–[5]. There are generally two methods to deal with the estimation problem in the presence of missing data: imputation and deletion [6]. If the positions of the missing data are known, then they can either be replaced by some values (deterministic or random) or the corresponding data can be removed from the dataset altogether. Removing data generally leads to performance degradation while data imputation results in bias estimates [2], [6], [7].

In this work, we examine how a connected network of agents, with each one of them subjected to a stream of data with incomplete regression information, can cooperate with each other to estimate the underlying model parameters in the presence of missing data. We explain how to adjust the traditional diffusion strategies through (de)regularization in order to eliminate the bias introduced by imputation. We consider two applications: one dealing with a mental health survey and the other dealing with a household consumption survey.

Notation. We use lowercase letters to denote vectors, uppercase letters for matrices, plain letters for deterministic variables, boldface letters for random variables. We use $\odot$ and $\otimes$ for the Hadamard and Kronecker products, respectively. Moreover, $\mathrm{diag}\{x_1,\ldots,x_N\}$ denotes a diagonal matrix with diagonal elements $x_1,\ldots,x_N$. We use $\mathrm{col}\{a,b\}$ to represent a column vector with entries a and b, while I_M denotes the $M\times M$ identity matrix.

II. PROBLEM STATEMENT

Consider a connected network with N agents. Each agent senses a stream of wide-sense stationary data $\{\boldsymbol{d}_k(i),\boldsymbol{u}_{k,i}\}$ that satisfy the linear regression model:

$$\boldsymbol{d}_k(i)=\boldsymbol{u}_{k,i}w^o+\boldsymbol{v}_k(i),\quad k=1,\ldots,N \tag{1}$$

where k is the node index and i is the time index. Moreover, the row vector $\boldsymbol{u}_{k,i}$ denotes a zero mean random process with covariance matrix $R_{u,k}=\mathbb{E}\,\boldsymbol{u}^*_{k,i}\boldsymbol{u}_{k,i}>0$, while $\boldsymbol{v}_k(i)$ is a zero mean white noise process with variance $\sigma^2_{v,k}$. The vector $w^o\in\mathbb{R}^M$ is the unknown parameter that the network is interested in estimating.

Assumption 1: We assume that the regression and noise processes are each spatially independent and temporally white. In addition, we assume that $\boldsymbol{u}_{\ell,i}$ and $\boldsymbol{v}_k(j)$ are independent of each other for all ℓ,i,k, and j. ∎

In this study, we examine the situation in which some entries in the regression vectors may be missing at random due to a variety of reasons, including incomplete information or censoring. We denote the incomplete regressor by $\bar{\boldsymbol{u}}_{k,i}$ and express it in the form:

$$\bar{\boldsymbol{u}}_{k,i}=\boldsymbol{u}_{k,i}(I_M-\boldsymbol{F}_{k,i})+\boldsymbol{\xi}_{k,i}\boldsymbol{F}_{k,i} \tag{2}$$

where $\boldsymbol{F}_{k,i}=\mathrm{diag}\{\boldsymbol{f}^1_{k,i},\ldots,\boldsymbol{f}^M_{k,i}\}$ consists of random indicator variables, $\boldsymbol{f}^j_{k,i}\in\{0,1\}$. Each variable $\boldsymbol{f}^j_{k,i}$ is equal to one with some probability p and equal to zero with probability $1-p$. The value of p represents the likelihood that the $j-$th entry of the regression vector $\boldsymbol{u}_{k,i}$ is missing at time i. In that case, the missing entry is assumed to be replaced by an entry from the zero mean perturbation variable $\boldsymbol{\xi}_{k,i}$.

Assumption 2: We assume that the random variables $\boldsymbol{u}_{k,i}$, $\boldsymbol{f}^j_{k,i}$, and $\boldsymbol{\xi}_{k,i}$ are independent of each other. We also assume that the random process $\boldsymbol{\xi}_{k,i}$ is temporally white and spatially independent with covariance matrix $\mathbb{E}\,\boldsymbol{\xi}^*_{k,i}\boldsymbol{\xi}_{k,i}=\sigma^2_\xi I_M$. ∎

From model (1), the minimum mean-square-error (MSE) estimate of the unknown vector w^o based on the data collected at node k is given by [8]:

$$w^o_k=R^{-1}_{u,k}\,r_{du,k} \tag{3}$$

where $r_{du,k}\triangleq\mathbb{E}\,\boldsymbol{d}_k(i)\boldsymbol{u}^*_{k,i}$. It is easy to verify from (1) that $w^o_k=w^o$ so that the MSE solution allows node k to recover the unknown w^o if the actual moments $\{R_{u,k},r_{du,k}\}$ happen to be known. The resulting mean-square-error is

$$J_{k,\min}\triangleq\sigma^2_{d,k}-r^*_{du,k}R^{-1}_{u,k}r_{du,k}=\sigma^2_{v,k}. \tag{4}$$

This work was performed while M. R. Gholami was a visiting student at the UCLA Adaptive Systems Laboratory. The work was supported in part by the the Swedish Research Council (contract no. 2007-6363) and in part by Ericsson's Research Foundation (Reference: FOSTIFT-12:021 and FOSTIFT-13:031). The work of A. H. Sayed was supported by NSF grant CCF-1011918.

Now, we investigate the MSE estimate that would result based on the censored regressor, $\bar{\boldsymbol{u}}_{k,i}$. We first introduce the following matrices, which will be used in the analysis:

$$
\begin{aligned}
P_1 &\triangleq (2p-p^2)\mathbb{1}_M\mathbb{1}_M^T-(p-p^2)I_M &(5)\\
P_2 &\triangleq p^2\mathbb{1}_M\mathbb{1}_M^T+(p-p^2)I_M &(6)\\
P &\triangleq pI_M &(7)
\end{aligned}
$$

where $\mathbb{1}_M$ is the M-column vector with all its entries equal to one. The censored estimate is given by

$$\bar{w}_k^o = R_{\bar{u},k}^{-1}\, r_{d\bar{u},k} \tag{8}$$

with the covariance matrix $R_{\bar{u},k}$ given by

$$R_{\bar{u},k} \triangleq R_{u,k}+R_{r,k} \tag{9}$$

where we are introducing

$$R_{r,k} \triangleq -P_1\odot R_{u,k}+p\sigma_\xi^2 I_M. \tag{10}$$

Moreover, the cross correlation vector $r_{d\bar{u},k}$ is given by

$$r_{d\bar{u},k} \triangleq \mathbb{E}\,\boldsymbol{d}_k\bar{\boldsymbol{u}}_{k,i}^* = (1-p)r_{du,k}. \tag{11}$$

We assume the matrix $(I_M+R_{u,k}^{-1}R_{r,k})$ is invertible. We can then relate $\bar{w}_k^o$ from (8) to w_k^o from (3) as follows:

$$\bar{w}_k^o = (I_M-Q_k)(1-p)w^o \tag{12}$$

where

$$Q_k \triangleq R_{u,k}^{-1}R_{r,k}(I_M+R_{u,k}^{-1}R_{r,k})^{-1}.$$

It is observed from (12) that the new estimate is biased relative to w^o.

III. Adaptive Distributed Strategy

To mitigate the bias, we associate the following individual cost with each agent k:

$$\bar{J}_k(w) \triangleq \mathbb{E}|\boldsymbol{d}_k(i)-\bar{\boldsymbol{u}}_{k,i}w|^2-\|w\|_{T_k}^2 \tag{13}$$

where T_k is a symmetric matrix to be chosen. The minimizer of (13) is seen to be

$$\bar{w}_k^o = (R_{\bar{u},k}-T_k)^{-1}(1-p)r_{du,k}. \tag{14}$$

Therefore, if we select

$$T_k = pR_{u,k}-P_1\odot R_{u,k}+p\sigma_\xi^2 I_M \tag{15}$$

then $\bar{w}_k^o = w_k^o$ (i.e., the (de)regularised estimate from (14) would coincide with the unbiased estimate from (3)). The corresponding mean-square-error for this choice of T_k is given by

$$
\begin{aligned}
\bar{J}_{k,\min} &\triangleq \sigma_{d,k}^2-(1-p)^2 r_{du,k}^*(R_{\bar{u},k}-T_k)^{-1}r_{du,k}\\
&= J_{k,\min}+p r_{du,k}^* R_{u,k}^{-1} r_{du,k} > J_{k,\min}. &(16)
\end{aligned}
$$

For the remainder of the paper, we introduce a simplifying assumption.

Assumption 3: The covariance matrix $R_{u,k}$ is diagonal, which is satisfied if the entries of the regression vector $\boldsymbol{u}_{k,i}$ are uncorrelated with each other. ■

Under Assumption 3, it can be verified that

$$
\begin{aligned}
R_{r,k} &= -pR_{u,k}+p\sigma_\xi^2 I_M &(17)\\
R_{\bar{u},k} &= (1-p)R_{u,k}+p\sigma_\xi^2 I_M &(18)\\
T_k &= p\sigma_\xi^2 I_M. &(19)
\end{aligned}
$$

We first assume that p and σ_ξ^2 are known. Later, we estimate σ_ξ^2 from the data, assuming an estimate for p is available.

To develop a distributed algorithm, we let $\mathcal{N}_k$ denote the set of neighbors of agent k. The network then seeks to solve:

$$\underset{w\in\mathbb{R}^M}{\text{minimize}} \;\sum_{k=1}^N \bar{J}_k(w). \tag{20}$$

Following arguments similar to [9], [10], we can motivate the following Adapt-then-Combine (ATC) diffusion strategy:

$$
\begin{aligned}
\boldsymbol{\phi}_{k,i} &= (1+\mu_k p\sigma_\xi^2)\boldsymbol{w}_{k,i-1}+\mu_k\bar{\boldsymbol{u}}_{k,i}^*[\boldsymbol{d}_k(i)-\bar{\boldsymbol{u}}_{k,i}\boldsymbol{w}_{k,i-1}]\\
\boldsymbol{w}_{k,i} &= \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\boldsymbol{\phi}_{\ell,i} &(21)
\end{aligned}
$$

where μ_k is a small step size and the convex combination coefficients $\{a_{\ell k}\}$ satisfy

$$\sum_{\ell\in\mathcal{N}_k} a_{\ell k}=1, \quad a_{\ell k}=0 \text{ for } \ell\notin\mathcal{N}_k. \tag{22}$$

IV. Performance Analysis

A. Error Dynamics

We associate with each agent the error vector

$$\tilde{\boldsymbol{w}}_{k,i} \triangleq w^o-\boldsymbol{w}_{k,i} \tag{23}$$

and collect the errors from across the network into the block vector:

$$\tilde{\boldsymbol{w}}_i \triangleq \text{col}\{\tilde{\boldsymbol{w}}_{1,i},\cdots,\tilde{\boldsymbol{w}}_{N,i}\}. \tag{24}$$

Then, it can be verified that $\tilde{\boldsymbol{w}}_i$ evolves according to the following dynamics:

$$
\begin{aligned}
\tilde{\boldsymbol{w}}_i =& \mathcal{A}^T[I_{NM}-\mathcal{M}(\bar{\boldsymbol{\mathcal{R}}}_i-p\sigma_\xi^2 I_{NM})]\tilde{\boldsymbol{w}}_{i-1}-\mathcal{A}^T\mathcal{M}\boldsymbol{s}_i\\
&-\mathcal{A}^T\mathcal{M}(\boldsymbol{\mathcal{R}}_{e,i}+p\sigma_\xi^2 I_{NM})w_e^o &(25)
\end{aligned}
$$

where

$$
\begin{aligned}
&w_e^o \triangleq \mathbb{1}_N\otimes w^o, \quad \mathcal{A}\triangleq A\otimes I_M &(26)\\
&\bar{\boldsymbol{\mathcal{R}}}_i \triangleq \text{diag}\{\bar{\boldsymbol{u}}_{1,i}^*\bar{\boldsymbol{u}}_{1,i},\bar{\boldsymbol{u}}_{2,i}^*\bar{\boldsymbol{u}}_{2,i},\ldots,\bar{\boldsymbol{u}}_{N,i}^*\bar{\boldsymbol{u}}_{N,i}\} &(27)\\
&\boldsymbol{\mathcal{R}}_{e,i} \triangleq \text{diag}\left\{\{\bar{\boldsymbol{u}}_{k,i}^*(\boldsymbol{u}_{k,i}-\boldsymbol{\xi}_{k,i})\boldsymbol{F}_{k,i}\}_{k=1,\ldots,N}\right\} &(28)\\
&\mathcal{M} \triangleq \text{diag}\{\mu_1 I_M,\mu_2 I_M,\ldots,\mu_N I_N\} &(29)\\
&\boldsymbol{s}_i \triangleq \text{col}\{\bar{\boldsymbol{u}}_{1,i}^*\boldsymbol{v}_1(i),\ldots,\bar{\boldsymbol{u}}_{N,i}^*\boldsymbol{v}_N(i)\} &(30)
\end{aligned}
$$

where the matrix A is left-stochastic, i.e., $A^T\mathbb{1}_M=\mathbb{1}_M$, with its (ℓ,k) entry equal to $a_{\ell k}$. From the above definitions, we get

$$
\begin{aligned}
\mathbb{E}\boldsymbol{s}_i &= 0 &(31)\\
\mathcal{S} &\triangleq \mathbb{E}\boldsymbol{s}_i\boldsymbol{s}_i^* = \text{diag}\{\sigma_{v,1}^2 R_{\bar{u},1},\ldots,\sigma_{v,N}^2 R_{\bar{u},N}\} &(32)\\
\bar{\mathcal{R}} &\triangleq \mathbb{E}\bar{\boldsymbol{\mathcal{R}}}_i = \text{diag}\{R_{\bar{u},1},\ldots,R_{\bar{u},N}\}\\
&= (1-p)\text{diag}\{R_{u,1},\ldots,R_{u,N}\}+p\sigma_\xi^2 I_{NM} &(33)\\
\mathbb{E}\boldsymbol{\mathcal{R}}_{e,i} &= -p\sigma_\xi^2 I_{NM}. &(34)
\end{aligned}
$$

B. Mean Stability Analysis

Since the variables $\boldsymbol{u}_{k,i}$ and $\boldsymbol{\xi}_{k,i}$ are temporally white and spatially independent, then the error vectors $\tilde{\boldsymbol{w}}_{\ell,j}$ are independent of $\boldsymbol{u}_{k,i}$ and $\boldsymbol{\xi}_{k,i}$ for all j if $k\neq\ell$ and for $k=\ell$ if $j\leq i-1$. Therefore, taking the expectation of both sides of (25), we get

$$\mathbb{E}\tilde{\boldsymbol{w}}_i = \mathcal{A}^T[I_{NM}-\mathcal{M}(\bar{\mathcal{R}}-p\sigma_\xi^2 I_{NM})]\mathbb{E}\tilde{\boldsymbol{w}}_{i-1}. \tag{35}$$

The recursion in (35) is stable if the step sizes are chosen to satisfy

$$0<\mu_k<\frac{2}{(1-p)\lambda_{\max}(R_{u,k})} \tag{36}$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of its matrix argument. It is seen that the estimator is asymptotically unbiased, i.e., $\lim_{i\to\infty}\mathbb{E}\tilde{\boldsymbol{w}}_i=0$.

C. Mean Variance Analysis

We rewrite (25) more compactly as

$$\tilde{\boldsymbol{w}}_i = \boldsymbol{\mathcal{B}}_i \tilde{\boldsymbol{w}}_{i-1} - \mathcal{G}\boldsymbol{s}_i - \boldsymbol{\mathcal{D}}_i w_e^o \tag{37}$$

where

$$\boldsymbol{\mathcal{B}}_i \triangleq \mathcal{A}^T[I_{NM} - \mathcal{M}(\bar{\boldsymbol{\mathcal{R}}}_i - p\sigma_\xi^2 I_{NM})] \tag{38}$$
$$\boldsymbol{\mathcal{D}}_i \triangleq \mathcal{A}^T\mathcal{M}(\boldsymbol{\mathcal{R}}_{e,i} + p\sigma_\xi^2 I_{NM}) \tag{39}$$
$$\mathcal{G} \triangleq \mathcal{A}^T\mathcal{M}. \tag{40}$$

The mean-square error analysis relies on evaluating a weighted variance of the error vector. Let Σ denote an arbitrary nonnegative definite matrix that we are free to choose. Considering $\lim_{i\to\infty}\mathbb{E}\tilde{\boldsymbol{w}}_i = 0$, $\mathbb{E}\boldsymbol{s}_i = 0$, and the independence between $\tilde{\boldsymbol{w}}_i$ and $\boldsymbol{s}_i$, from (37) we get

$$\lim_{i\to\infty}\mathbb{E}\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2 = \lim_{i\to\infty}\big[\mathbb{E}(\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{\mathcal{B}}_i^*\Sigma\boldsymbol{\mathcal{B}}_i\tilde{\boldsymbol{w}}_{i-1}) + \mathbb{E}(\boldsymbol{s}_i^*\mathcal{G}^T\Sigma\mathcal{G}\boldsymbol{s}_i) + \mathbb{E}(w_e^{o*}\boldsymbol{\mathcal{D}}_i^*\Sigma\boldsymbol{\mathcal{D}}_i w_e^o)\big] \tag{41}$$

from which we can write

$$\lim_{i\to\infty}\mathbb{E}\|\tilde{\boldsymbol{w}}_i\|_\sigma^2 = \lim_{i\to\infty}\mathbb{E}\|\tilde{\boldsymbol{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 + [\mathrm{vec}(\mathcal{Z}^T + \mathcal{Y}^T)]^T\sigma \tag{42}$$

in terms of the following quantities:

$$\mathcal{F} \triangleq \mathcal{A}\otimes\mathcal{A} - \mathcal{A}\otimes(\bar{\mathcal{R}} - p\sigma_\xi^2 I_{NM})^T\mathcal{M}\mathcal{A}^T - (\bar{\mathcal{R}} - p\sigma_\xi^2 I_{NM})^T\mathcal{M}\mathcal{A}\otimes\mathcal{A} + O(\mathcal{M}^2) \tag{43}$$
$$\mathcal{Z} \triangleq -p^2\sigma_\xi^4\mathcal{A}^T\mathcal{M}w_e^o w_e^{o*}\mathcal{M}\mathcal{A} + \mathcal{A}^T\mathcal{M}\mathbb{E}(\boldsymbol{\mathcal{R}}_{e,i}w_e^o w_e^{o*}\boldsymbol{\mathcal{R}}_{e,i}^*)\mathcal{M}\mathcal{A} \tag{44}$$
$$\mathcal{Y} \triangleq \mathcal{G}\mathcal{S}\mathcal{G}^T. \tag{45}$$

The shorthand notation σ in (42) represents the weighting matrix Σ and is given by $\sigma = \mathrm{vec}(\Sigma)$, where the vec operator vectorizes a matrix by placing its columns on top of each other. Following arguments similar to [9], the diffusion algorithm in (25) can be verified to be stable in the mean-square-error if the matrix $\mathcal{F}$ is stable, which can be satisfied for sufficiently small step-sizes.

We can then evaluate the network and individual MSDs as

$$\mathrm{MSD}^{\mathrm{network}} \triangleq \lim_{i\to\infty}\frac{1}{N}\sum_{k=1}^N\mathbb{E}\|\tilde{\boldsymbol{w}}_{k,i}\|^2 = \lim_{i\to\infty}\mathbb{E}\|\tilde{\boldsymbol{w}}_i\|_{\frac{1}{N}}^2 = \frac{1}{N}[\mathrm{vec}(\mathcal{Z}^T + \mathcal{Y}^T)]^T(I-\mathcal{F})^{-1}\mathrm{vec}(I_{NM}) \tag{46}$$
$$\mathrm{MSD}_k \triangleq \lim_{i\to\infty}\mathbb{E}\|\tilde{\boldsymbol{w}}_{k,i}\|^2 = \lim_{i\to\infty}\mathbb{E}\|\tilde{\boldsymbol{w}}_i\|_{\mathcal{I}_k}^2 = [\mathrm{vec}(\mathcal{Z}^T + \mathcal{Y}^T)]^T(I-\mathcal{F})^{-1}\mathrm{vec}(\mathcal{I}_k) \tag{47}$$

where $\mathcal{I}_k \triangleq \mathrm{diag}\{0,\ldots,0,I_M,0,\ldots,0\}$, with the identity matrix appearing in the $k-$th block location.

D. Estimation of Regularization Parameter

In the sequel, we suggest one way to estimate the (de)regularization coefficient σ_ξ^2 from the data. For a small probability of missing p, we have $P_2 \simeq pI_M$. Moreover, note that

$$J_{k,\min} \triangleq \sigma_{v,k}^2 = \mathbb{E}|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}w^o|^2 = \mathbb{E}|\boldsymbol{d}_k(i) - \bar{\boldsymbol{u}}_{k,i}w^o|^2 - p\|w^o\|_{R_{u,k}}^2 - p\sigma_\xi^2\|w^o\|^2. \tag{48}$$

After a sufficient number of iterations, the estimate $\boldsymbol{w}_{k,i}$ in (21) will get close to w^o. We replace the optimal w^o by $\boldsymbol{w}_{k,i}$ and get

$$J_{k,\min} = \sigma_{v,k}^2 \simeq \mathbb{E}|\boldsymbol{e}_k(i)|^2 - p\|\boldsymbol{w}_{k,i}\|_{R_{u,k}}^2 - p\sigma_\xi^2\|\boldsymbol{w}_{k,i}\|^2 \tag{49}$$

where $\boldsymbol{e}_k(i) \triangleq \boldsymbol{d}_k(i) - \bar{\boldsymbol{u}}_{k,i}\boldsymbol{w}_{k,i}$. It is still not possible to estimate $p\sigma_\xi^2$ directly from (49). We need to assume that either the variance of the original regressor or the probability of missing is known. Suppose that an estimate for p is available, say, $\hat{p}$. From (18), we can write

$$R_{u,k} \simeq \frac{1}{1-\hat{p}}R_{\bar{u},k} - \frac{\hat{p}}{1-\hat{p}}\sigma_\xi^2 I_M. \tag{50}$$

Substituting $R_{u,k}$ in (49) by the right-hand side of (50), we can estimate the variance σ_ξ^2 at node k as follows:

$$\hat{\sigma}_{\xi,k}^2 \overset{(a)}{\simeq} \frac{(1-\hat{p})\mathbb{E}|\boldsymbol{e}_k(i)|^2 - \hat{p}\|\boldsymbol{w}_{k,i}\|_{\boldsymbol{R}_{\bar{u},k}}^2}{\hat{p}(1-2\hat{p})\|\boldsymbol{w}_{k,i}\|^2} \tag{51}$$

where in (a) we assumed that the noise variance $\sigma_{v,k}^2$ is sufficiently small compared to other terms. Since $\mathbb{E}|\boldsymbol{e}_k(i)|^2$ and the diagonal matrix $R_{\bar{u},k}$ are unknown, we estimate them by means of the following smoothing filters from data realizations:

$$\hat{\boldsymbol{R}}_{\bar{u},k}(i) = (1-\alpha_1)\hat{\boldsymbol{R}}_{\bar{u},k}(i-1) + \alpha_1(\bar{\boldsymbol{u}}_{k,i}^*\bar{\boldsymbol{u}}_{k,i})\odot I_M \tag{52}$$
$$\boldsymbol{f}_k(i) = (1-\alpha_2)\boldsymbol{f}_k(i-1) + \alpha_2|\boldsymbol{e}_k(i)|^2 \tag{53}$$
$$\boldsymbol{g}(i) = \frac{(1-\hat{p})\boldsymbol{f}_k(i) - \hat{p}\|\boldsymbol{w}_{k,i}\|_{\hat{\boldsymbol{R}}_{\bar{u},k}(i)}^2}{\hat{p}(1-2\hat{p})\|\boldsymbol{w}_{k,i}\|^2} \tag{54}$$
$$\hat{\boldsymbol{\sigma}}_{\xi,k}^2(i) = (1-\alpha_3)\hat{\boldsymbol{\sigma}}_{\xi,k}^2(i-1) + \alpha_3\boldsymbol{g}(i) \tag{55}$$

where $0 < \alpha_i \ll 1$, for $i = 1,2,3$. To prevent large fluctuations in estimating $\hat{\boldsymbol{\sigma}}_{\xi,k}^2(i)$, we used a smoothing filter for updating $\hat{\boldsymbol{\sigma}}_{\xi,k}^2(i)$ in (52).

V. SIMULATION RESULTS

In this section, two applications are considered. In the simulations, we consider the connected network of 7 agents shown in Fig. 1 and employ the uniform combination rule $a_{\ell,k} = 1/|\mathcal{N}_k|$.

A. Household Consumption

Household consumption depends on a number of parameters such as income, wealth, family size, and retirement status [2]. We consider the following log-form model for household consumption [2]:

$$\ln\boldsymbol{c}_\ell(i) = \alpha + (\ln\boldsymbol{l}_{\ell,i})\beta_1 + (\ln\boldsymbol{m}_{\ell,i}^p)\beta_2 + (\ln\boldsymbol{m}_{\ell,i}^c)\beta_3 + \boldsymbol{t}_{\ell,i}\beta_4 + \boldsymbol{\epsilon}_\ell(i)$$

where $\boldsymbol{c}_\ell(i)$ is the consumption of household ℓ at time i, $\boldsymbol{l}_{\ell,i}$ is total wealth, which is assumed to be censored, $\boldsymbol{m}_{\ell,i}^p$ is the permanent part of income, $\boldsymbol{m}_{\ell,i}^c$ is the current income, $\boldsymbol{t}_{\ell,i}$ refers to the retirement status and family size. The modeling error $\boldsymbol{\epsilon}_\ell(i)$ is assumed to be zero mean. Similar to [2], we only consider the first 4 components of the regressor, i.e., $\beta_4 = 0$. If we subtract the mean of the measurement from both sides, we arrive at the model

$$\boldsymbol{d}_\ell^c(i) = \boldsymbol{u}_{\ell,i}^c w_c + \boldsymbol{\epsilon}_\ell(i) \tag{56}$$

where

$$w_c = [\beta_1\ \beta_2\ \beta_3]^T, \qquad \boldsymbol{d}_\ell^c(i) \triangleq \ln\boldsymbol{c}_\ell(i) - \mathbb{E}\left(\ln\boldsymbol{c}_\ell(i)\right)$$
$$\boldsymbol{u}_{\ell,i}^c \triangleq [\ln\boldsymbol{l}_{\ell,i}\ \ \ln\boldsymbol{m}_{\ell,i}^p\ \ \ln\boldsymbol{m}_{\ell,i}^c] - \mathbb{E}\,[\ln\boldsymbol{l}_{\ell,i}\ \ \ln\boldsymbol{m}_{\ell,i}^p\ \ \ln\boldsymbol{m}_{\ell,i}^c].$$

For the complete data, the unknown parameters can be estimated via a least-squares procedure to yield $\hat{w}_c = [0.054\ 0.182\ 0.24]^T$. We generate data according to $\hat{w}_c$ and assume that the regressor $\boldsymbol{u}_{\ell,i}^c$ has Gaussian distribution. We model $\boldsymbol{\epsilon}_\ell(i)$ by a zero mean Gaussian random variable. We further assume that the log of wealth is randomly missed and we consider a uniform distribution over $[-1,1]$ for the missing variable, thus $\sigma_\xi^2 = 2/3$. In the survey, it has been observed that approximately 30% of total wealth, including housing and stock market, are censored [2]. We use $\alpha_1 = \alpha_2 = \alpha_3 = 0.01$. In the simulation, we use $\mu_k = 0.01$. Fig. 2 shows the bias and the MSD of the estimator (52) for $p = 0.3$ at node 1. The results

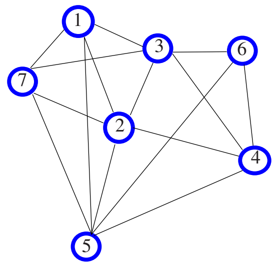

Fig. 1. The topology of the network.

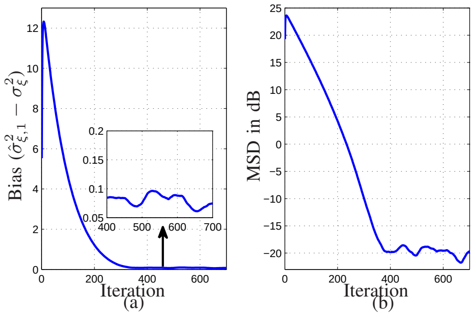

Fig. 2. Estimating the variance of missing noise, σ^2_{ξ}, from the data at node 1. (a) The bias of the estimate and (b) the MSD (in dB) of the estimate.

show that the algorithm is able to estimate the variance σ^2_{ξ} after a number of iterations. Next, we compare the performance of different algorithms. The results are shown in Fig. 3 for the modified diffusion (21) (MATC), the regular diffusion (ATC) [9], and for noncooperative (ncoop) behavior.

B. Mental Health Survey

We consider the following model, motivated by a mental health survey study run by various companies [11], [12]:

$$\boldsymbol{d}_k(i) = \boldsymbol{x}_{k,i}\theta + \boldsymbol{v}_k(i), \ i = 1, 2, \ldots, N, \tag{57}$$

where $\boldsymbol{d}_k(i)$ is the square root of total depression score for every individual i, $\boldsymbol{x}_{k,i} \in \mathbb{R}^6$ denotes the regressor (covariate) for every individual i, and $\boldsymbol{v}_k(i)$ is the modeling error. Index k refers to the company index and i is used for participant's index. The elements of $\boldsymbol{x}_{k,i}$ are defined in [11] and include variables such as income, age, and martial status. We apply the least squares technique to a subset of the data in [11] to find the estimate of $\hat{\theta} = [0.6352 \ \ -0.0230 \ 0.0163 \ 0.1993 \ 1.4157 \ \ -0.2367 \ 0.2419]^T$. Then, we generate simulated depression scores according to $\hat{\theta}$ and randomly realized regressors. Note that we generate (uniformly) zero mean random regressors $\boldsymbol{x}_{k,i}$. We further assume that the income

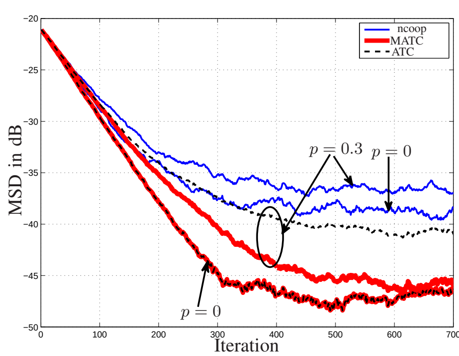

Fig. 3. The MSD of different algorithms.

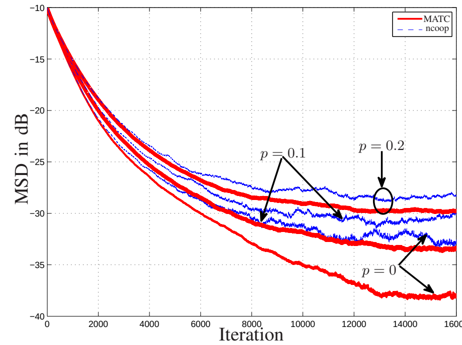

Fig. 4. MSD for modified ATC and noncooperative approach for different missing probabilities.

is missed with probability p in the simulation study. We consider a zero mean Gaussian distribution with standard deviation 0.05 for missing parts. The variance of measurement noise $\boldsymbol{v}_k(i)$, i.e., $\sigma^2_{v,k}$, is set to 0.05. In the simulation, we use $\mu_k = 0.008$. The MSDs are obtained by averaging over 40 experiments. In Fig. 4, we compare the performance of MATC diffusion and non-cooperative (ncoop) for different missing probabilities.

VI. CONCLUSIONS

In this paper, we have examined the estimation of an unknown vector over a connected network of agents, with each agent subjected to a stream of data with incomplete regressors. We have shown that the estimator in general is biased; hence, we have modified the cost function by a (de)regularisation term to mitigate the bias and obtained a distributed approach based on diffusion adaptation techniques. We have studied the mean-stability and performance of the proposed algorithm. We have also suggested a technique to estimate the (de)regularisation term from the data. We have evaluated the proposed algorithm for two applications in mental health and household consumption surveys.

REFERENCES

[1] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[2] R. Rigobon and T. M. Stoker, "Estimation with censored regressors: Basic issues," *International Economic Review*, vol. 48, no. 4, pp. 1441–1467, 2007.

[3] D. R. Cox and D. Oakes, *Analysis of Survival Data*. Chapman & Hall/CRC, 1984.

[4] C. K. Enders, *Applied Missing Data Analysis*. The Guilford Press, 2010.

[5] B. Marlin, R. S. Zemel, S. Roweis, and M. Slaney, "Collaborative filtering and the missing at random assumption," *arXiv preprint arXiv:1206.5267*, 2012.

[6] P. Royston, "Multiple imputation of missing values," *Stata Journal*, vol. 4, pp. 227–241, 2004.

[7] R. Rigobon and T. M. Stoker, "Bias from censored regressors," *Journal of Business & Economic Statistics*, vol. 27, no. 3, pp. 340–353, 2009.

[8] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall New Jersey, 2000, vol. 1.

[9] A. H. Sayed, "Diffusion adaptation over networks," *in E-Reference Signal Processing*, R. Chellapa and S. Theodoridis, editors, Elsevier, 2013. Also available as arXiv:1205.4220 [cs.MA], May 2012.

[10] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic, "Diffusion strategies for adaptation and learning," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 155–171, May 2013.

[11] A. A. Afifi and V. Clark, *Computer-Aided Multivariate Analysis*. CRC Press, 2004.

[12] R. J. Little, "Regression with missing x's: A review," *Journal of the American Statistical Association*, vol. 87, no. 420, pp. 1227–1237, 1992.